Stat 2 coursework

Alberto Plebani
ap2387

Number of words:

1

The trigonometric relationship between the four parameters θ, α, β, and x is presented in Equation (1). This relationship comes from the fact that we have a right triangle, where $x - \alpha$ and β are the cathetus.

$$\theta = \arctan\left(\frac{\beta}{x-\alpha}\right) \tag{1}$$

2

The normalised PDF for θ is presented in Equation (2), meaning θ is uniformly distributed between $-\pi/2$ and $+\pi/2$.

$$P_\theta(\theta) = \begin{cases} \frac{1}{\pi}\ if\ \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ 0\ otherwise \end{cases} \tag{2}$$

Knowing that $P_\theta(\theta)d\theta = P_x(x)dx$, we can get the PDF for x as $P_x(x) = P_\theta(\theta)\frac{d\theta}{dx}$. Using the relation in Equation 1, we know that $d\theta/dx$ is what is represented in Equation (3). Now, using again Equation (1), we can substitute $\tan\theta$ as $\beta/(x-\alpha)$, and by multiplying for the PDF $P_\theta(\theta)$ (Equation (2)), we get Equation (4). This Likelihood is a Lorentzian distribution, peaking at $x = \alpha$.

$$\frac{d\theta}{dx} = \frac{1}{\beta(1+\tan^2(\theta))} \tag{3}$$

$$\mathcal{L}(x|\alpha,\beta) = \frac{1}{\pi}\frac{\beta}{\beta^2 + (x-\alpha)^2} \tag{4}$$

3

Because the Likelihood is a Lorentzian centred around $x = \alpha$, the most likely location for any flash to be received is α. We can prove this by using the MLE, Because $d\mathcal{L}/dx =$

$-2(x-\alpha) \cdot \mathcal{L}$, thus having a maximum at $x = \alpha$.

The mean however, is not a good estimator. If we evaluate the expected value $\langle x \rangle$ we get an integral which is not defined, as displayed by Equation (5) (in line 2 I used $y = x - \alpha$). The second integral is convergent and it is equal to π, but the first integral is not defined because we have the difference between two logarithmic divergences.

$$\begin{aligned}
\langle x \rangle = \int x P_x(x) dx \simeq \int_{-\infty}^{+\infty} \frac{x\beta}{\beta^2 + (x-\alpha)^2} = \\
= \beta \int_{-\infty}^{+\infty} \frac{y+\alpha}{\beta^2 + y^2} = \\
= \beta \int_{-\infty}^{+\infty} \frac{y}{\beta^2 + y^2} + \beta\alpha \int_{-\infty}^{+\infty} \frac{1}{\beta^2 + y^2} = \\
= \left[\frac{\beta}{2} \ln(\beta^2 + y^2) \right]_{-\infty}^{+\infty} + \alpha \left[\arctan\left(\frac{y}{\beta} \right) \right]_{-\infty}^{+\infty}
\end{aligned} \tag{5}$$

In a similar way we can prove that the variance is infinite, because we have a term like $\int_{-\infty}^{+\infty} 1 dy$, which gives ∞, in addition to the same logarithmic divergence we saw above. Therefore, the mean is not a good estimator for this problem.

4

Because we have no prior knowledge on the location of the lighthouse, I decided to use uniform priors for both parameters. In the $(\alpha \times \beta)$ space, I defined a rectangle (whose boundaries can be changed when running the code by command line options), and the prior for each parameters is then $1/(b-a)$, where b and a are the upper and lower bound, respectively. The only contraint on the rectangle is that β is positive. For my final version of the code, I used the following ranges:

- $\alpha \in [-2, 2]$
- $\beta \in [0.1, 4]$

5

Using 20 flash locations from the lighthouse, I drew stochastic samples from the posterior distribution using a Metropolis-Hastings[1,2] (MH) Markov Chain Monte Carlo (MCMC).

This algorithm works in the following way:

[1] Metropolis at al. (1953) "Equation of State Calculations by Fast Computing Machines", Journal of Chemical Physics, **21** 6 1087-1092, doi:10.1063/1.1699114

[2] Hastings (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications", Biometrika, **57** 1, 97-109, doi:10.1093/biomet/57.1.97

- Define a target distribution $P(x)$, which is the distribution we want to sample from. In our case this is the product of the Likelihood in Equation (4) and the priors for α and β.
- Define a proposal distribution $Q(y|x)$. This distribution should be a distribution from which it is easy to draw samples from, and it is used to propose possibilities for the next steps. In my case I used a multivariate normal distribution
- Initialise the chain: set an initial state for the parameters ($x_0 = \alpha$)
- Now in iterative steps
  - *i)* Draw a proposed point y from the proposal distribution $Q(y|x_i)$
  - *ii)* Evaluate the MH acceptance probability $a = (P(y)Q(x_i|y))/(P(x_i)Q(y|x_i))$. In practise, you evaluate the log-likelihood at the points y and x_i, and you evaluate the difference
  - *iii)* Draw a number u from a uniform distribution between 0 and 1
  - *iv)* If $ln(u) < ln(a)$, accept the new point in the chain, thus setting x_{i+1} to y. If not, reject the point, and have $x_{i+1} = x_i$
- These steps are done iterativelly until the maximum number of samples (N, in my case 500000) is reached. This generates a Markov Chain with N elements, which are drawn from the target distribution, because as the chain goes on, the chain will reach a convergence
- Additionally, a burn-in is implemented, discarding the first N_b points (in my case 1000) which helps removing the dependence from the starting points.

The algorithm is implemented in the `HelpersFunctions.py` file in the `Helpers` folder. It uses the `multivariate_normal`[3] function from `scipy.stats` [4] package as proposal distribution.

5.1 Results

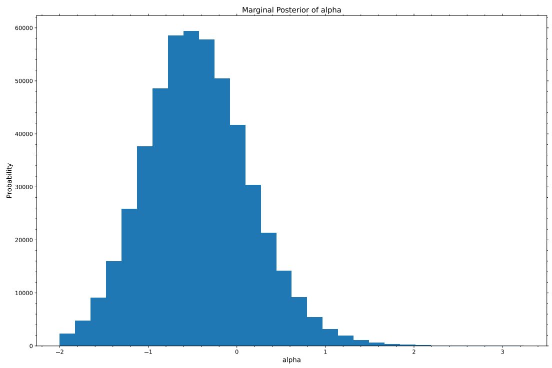
I ran the code with the following parameters:

- Number of samples generated: 500000
- Burn-in: 1000
- Range for α (uniform prior): $[-2, 2]$
- Range for β (uniform prior): $[0.1, 4]$
- Starting point: $(\alpha, \beta) = (0.3, 0.3)$

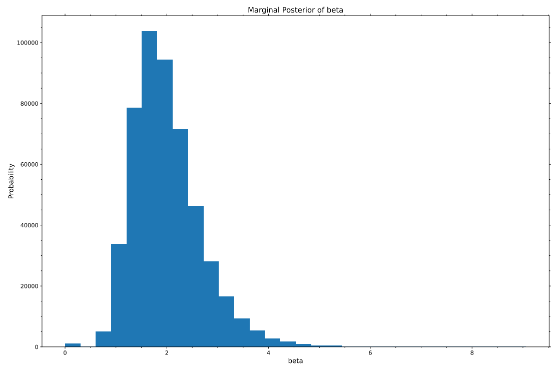
[3] Multivariate-normal
[4] scipy.stats

Parameter	Mean	Standard Deviation	Relative uncertainty
α	-0.4398	0.5944	1.35
β	1.998	0.6732	0.34

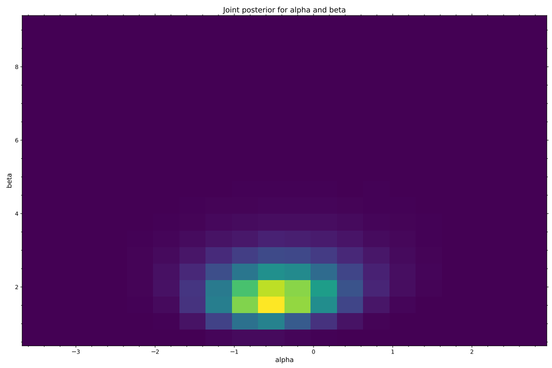
Table 1: Results for α and β



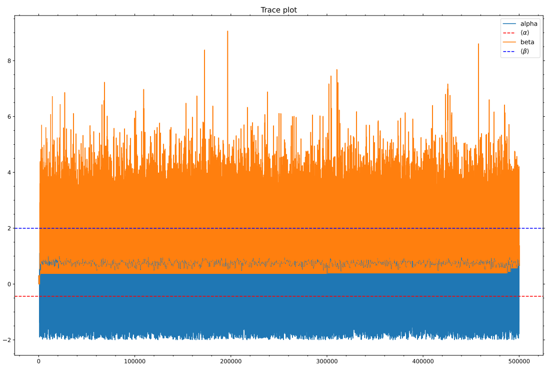
(a) Posterior for α



(b) Posterior for β



(c) Joint posterior for α and β



(d) Trace plot for α and β

Figure 1: Summary plots for the two parameters

Of the 500000 points generated, 146730 were accepted (acceptance=29%). The results are displayed in Table 1. In Figure 1 we can see the marginal PDFs for α and β (top), the joint PDF (bottom left) and the trace plot (bottom right), which ensures the convergence of the algorithm. From the latter, we can see that after a few iterations after the burn-in, all the points are sampled close to the true value.

6

For this section, we are considering that in addition to measuring the location, we have also measurements of the intensity of the lighthouse. The Likelihood for the intensity is presented in Equation (6). Because intensity and location measurements are independent, the total likelihood is the product of the two.

$$\mathcal{L}_I(\log I|\alpha,\beta,I_0) = \frac{\exp\left(\frac{-\left(\log I - \log(I_0)+\log(d^2)\right)}{2\sigma^2}\right)^2}{\sqrt{2\pi\sigma^2}} \tag{6}$$

Because the intensity measurements follow a log-normal distribution with an uncertainty $\sigma = 1$, I used a log-normal prior distribution for I_0, setting the mean value of the normal distribution to $\log(1.1)$, with this option that can be changed by command line option.

7

I repeated the same steps I used for part 5, including the intensity measurements, in addition to the position measurements. This time, the proposal distribution is a 3-dimensional normal distribution instead of the 2-dimensional one we used in the previous part. The rest of the problem is the same, with the Likelihood used to evaluate the proposal and the current positions which is the product of Equation (6) and (4).

7.1 Results

I ran the code with the following parameters:

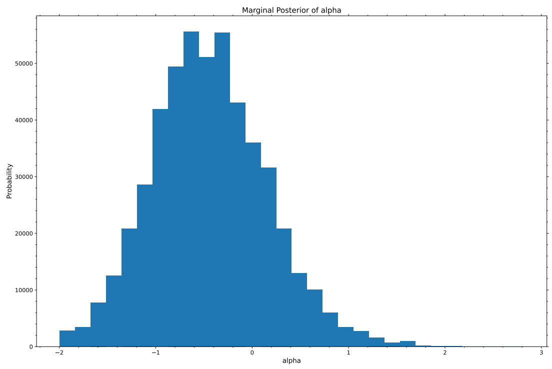
- Number of samples generated: 500000
- Burn-in: 1000
- Range for α (uniform prior): $[-2, 2]$
- Range for β (uniform prior): $[0.1, 4]$
- Mean of the log-normal distribution for I_0 : $\log(1.1)$
- Starting point: $(\alpha, \beta, I_0) = (0.3, 0.3, 0.3)$

Of the 500000 points generated, 12202 were accepted (acceptance=2.4%). The results are displayed in Table 2. We can see how the measurement of α is slighlty more precise now that we have added also measurements of the intensity. This was expected because the intensity depends on the distance from the source. Therefore, every measurement of the intensity allows us to gain additional information also on the position.

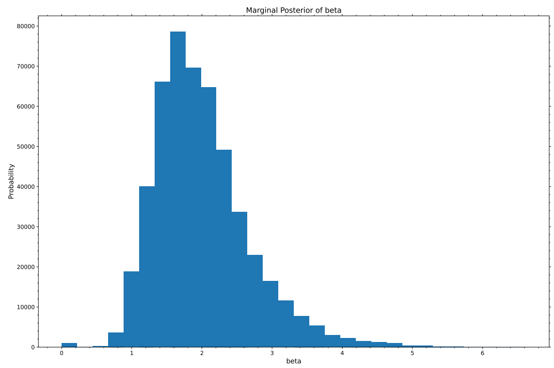
In Figure 2 we can see the marginal distributions for the three parameters, the 2D joint distributions for each pair of parameters, and finally the trace plot for all three parameters. We can see how once again the convergence is reached for all three parameters.

Finally, I present a couple of additional tests I performed.

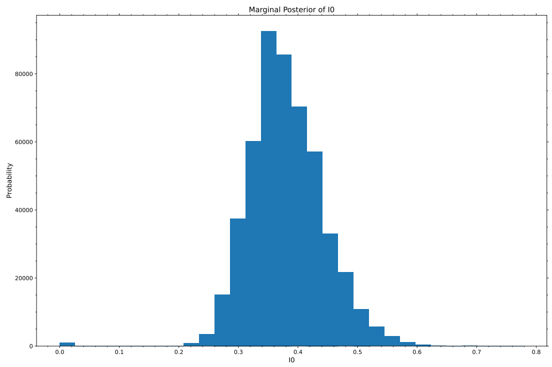
First of all, I tried to change the range in which the α and β priors are defined. I tried using a range of $\alpha \in [-5, 2]$ and $\beta \in [1, 3]$. The results changed slightly, but they remained



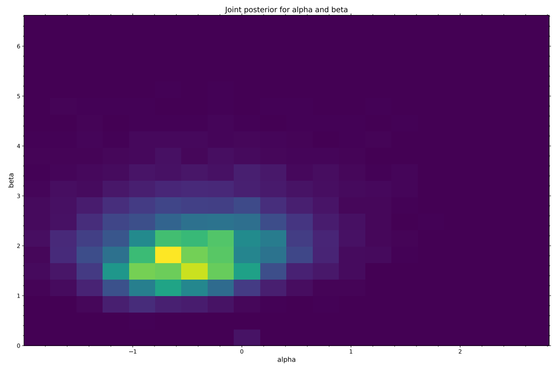
(a) Posterior for α



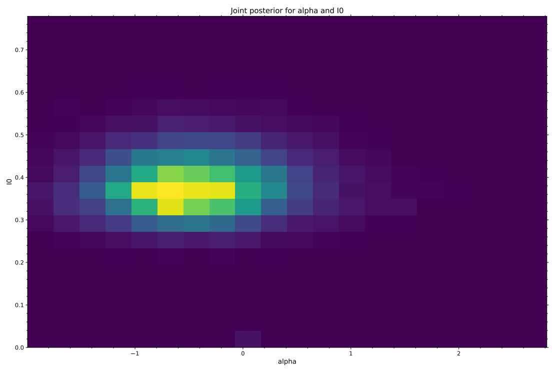
(b) Posterior for β



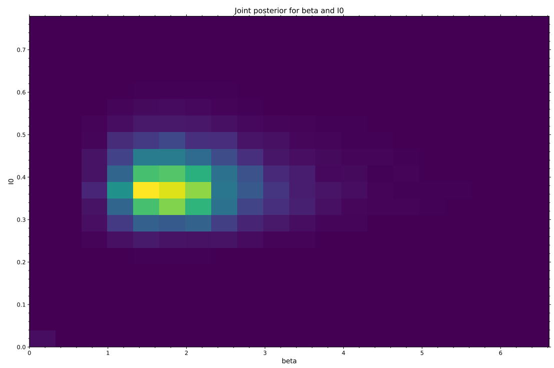
(c) Posterior for I_0



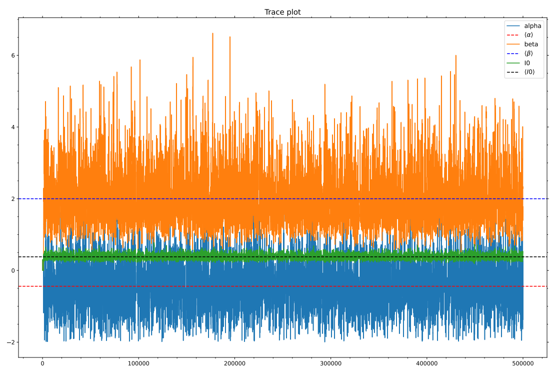
(d) Joint posterior for α and β



(e) Joint posterior for α and I_0



(f) Joint posterior for I_0 and β



(g) Trace plot for α, β, and I_0

Figure 2: Summary plots for the three parameters

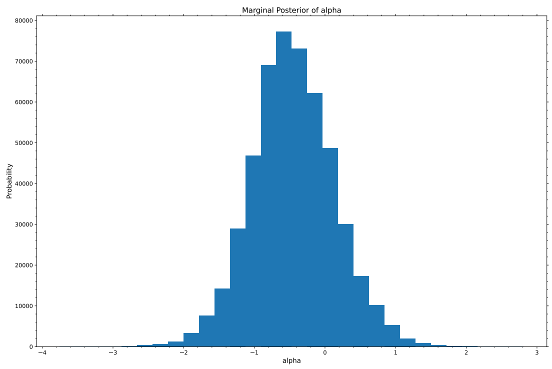
Parameter	Mean	Standard Deviation	Relative uncertainty
α	-0.440	0.594	1.34
β	1.999	0.681	0.34
I_0	0.380	0.062	0.16

Table 2: Results for α and β

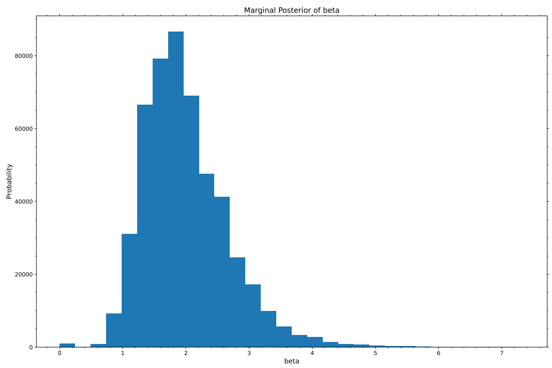
consistent with the previous measurement ($\alpha = -0.445 \pm 0.061$, $\beta = 2.00 \pm 0.68$ when using only the position; $\alpha = -0.453 \pm 0.586$, $\beta = 1.994 \pm 0.667$ and $I_0 = 0.380 \pm 0.063$). The plots are presented in Figure 3.

I also tried to change the starting points, to $(\alpha, \beta, I_0) = (1, 1, 1)$. We can see from Figure 4 that the trace plot is identical to the one obtained with the different starting point. This proves that the burn-in worked properly, because it ensured that the starting point does not matter in the outcome of the markov chain. Once again, the results don't change much with respect to the previous case.

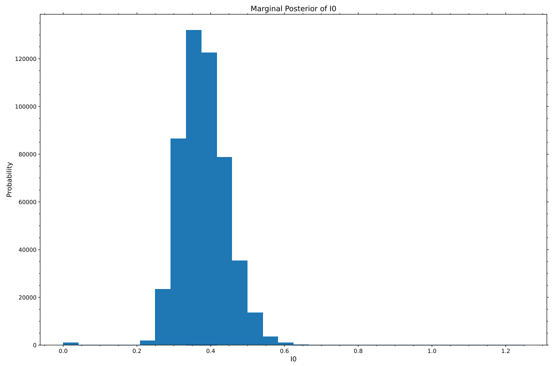
I also tried to change the mean value of I_0 and the covariance matrix used for the multivariate normal distribution, but the results remained the same.



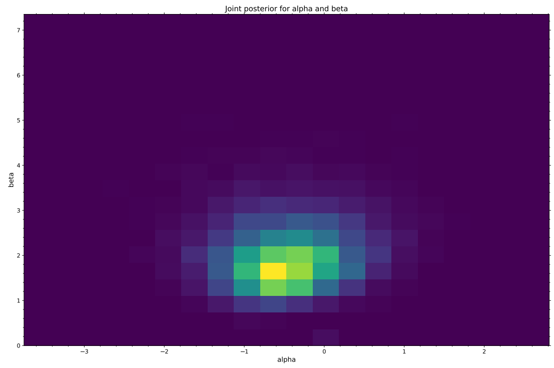
(a) Posterior for α



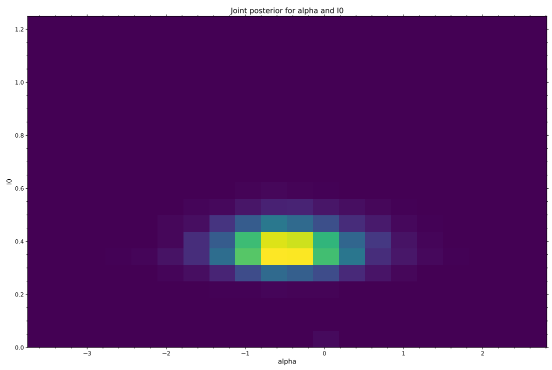
(b) Posterior for β



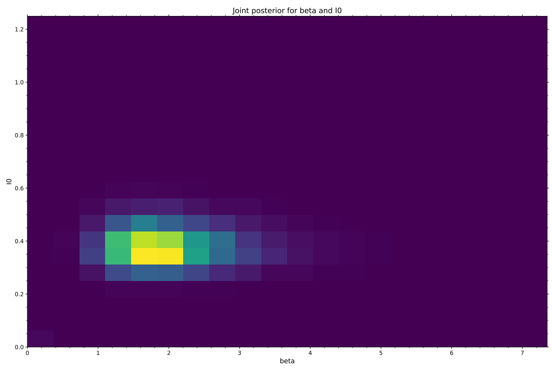
(c) Posterior for I_0



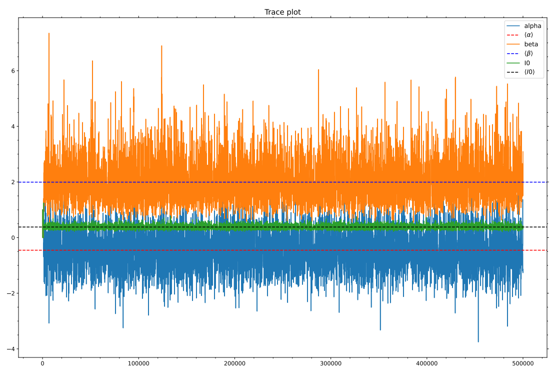
(d) Joint posterior for α and β



(e) Joint posterior for α and I_0

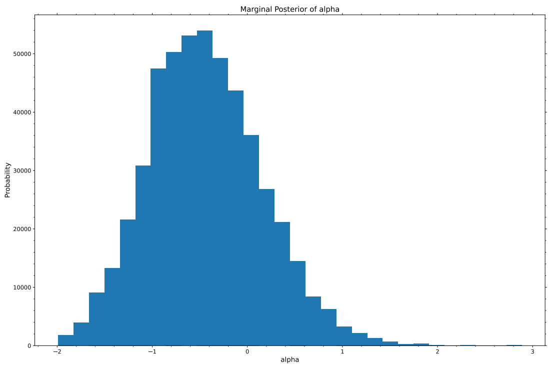


(f) Joint posterior for I_0 and β

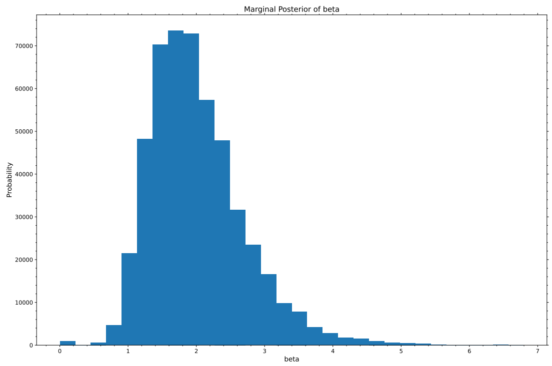


(g) Trace plot for α, β, and I_0

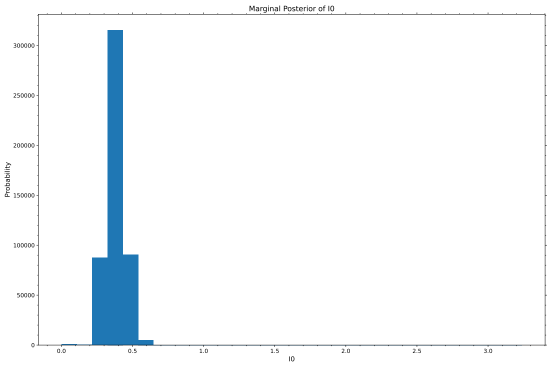
Figure 3: Summary plots for the three parameters, changing range of α and β



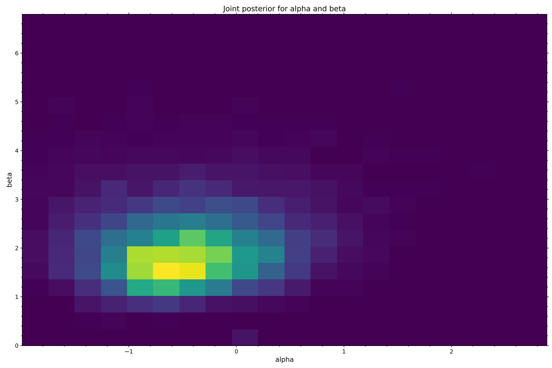
(a) Posterior for α



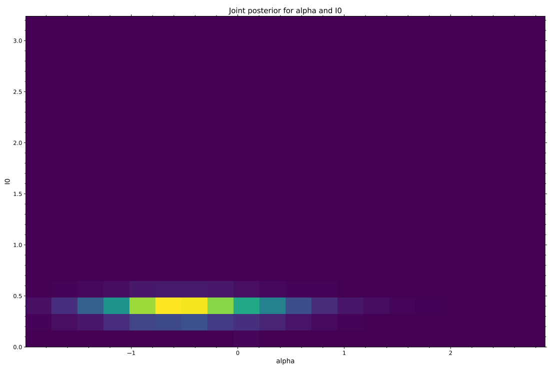
(b) Posterior for β



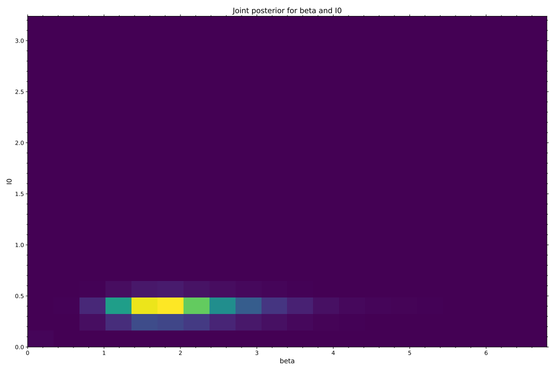
(c) Posterior for I_0



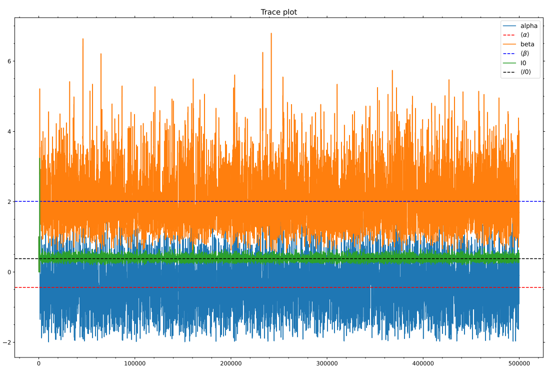
(d) Joint posterior for α and β



(e) Joint posterior for α and I_0



(f) Joint posterior for I_0 and β



(g) Trace plot for α, β, and I_0

Figure 4: Summary plots for the three parameters, changing the starting point